

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 1 016 076 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:

02.05.2003 Bulletin 2003/18

(51) Int Cl.7: **G10L 15/26**

(86) International application number:

PCT/US98/19433

(21) Application number: **98947098.4**

(87) International publication number:

WO 99/014743 (25.03.1999 Gazette 1999/12)

(22) Date of filing: **17.09.1998**

**(54) SYSTEM AND METHOD USING NATURAL LANGUAGE UNDERSTANDING FOR SPEECH
CONTROLLED APPLICATION**

**NATÜRLICHE SPRACHE VERSTEHENDES VERFAHREN UND VERSTEHENDE VORRICHTUNG
ZUR SPRACHSTEUERUNG EINER ANWENDUNG**

**SYSTEME ET PROCEDE UTILISANT LA COMPREHENSION D'UN LANGAGE NATUREL AFIN DE
L'APPLIQUER A UNE COMMANDE VOCALE**

(84) Designated Contracting States:

**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE**

• **DIEDRICHS, Raymond, Alan**

Chalfont, PA 18194 (US)

• **WALSH, Joseph, Patrick**

Spring City, PA 19475 (US)

(30) Priority: **17.09.1997 US 932938**

(43) Date of publication of application:

05.07.2000 Bulletin 2000/27

(74) Representative: **Modiano, Guido, Dr.-Ing. et al**

Modiano, Josif, Pisanty & Staub,

Baaderstrasse 3

80469 München (DE)

(73) Proprietor: **UNISYS CORPORATION**

Blue Bell, PA 19424-0001 (US)

(56) References cited:

WO-A-97/10589

US-A- 5 390 279

(72) Inventors:

• **SCHOLZ, Karl, Wilmer**

Strafford, PA 19087 (US)

• **BLUE, Reginald, Victor**

Exton, PA 19341 (US)

• **PATENT ABSTRACTS OF JAPAN vol. 096, no.
011, 29 November 1996 & JP 08 194600 A (NEC
CORP), 30 July 1996**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Description**Field of the Invention**

[0001] This invention relates generally to computerized natural language systems. More particularly, it relates to a computer system and method for providing speech understanding capabilities to an interactive voice response system. It further relates to a computer system and method for interpreting spoken utterances in a constrained speech recognition application.

Description of the Related Art

[0002] Computers have become a mainstay in our everyday lives. Many of us spend hours a day using the machines at work, home and even while shopping. Using a computer, however, has always been on the machine's terms. A mouse, pushbuttons and keyboards have always been somewhat of an unnatural way to tell the computers what we want. However, as computer technology continues to advance, the computer is edging towards communicating with humans on our terms: the spoken word.

[0003] There are essentially two steps in creating a computer that can speak with humans. First, the computer needs an automatic speech recognition to detect the spoken words and convert them into some form of computer-readable data, such as simple text. Second, the computer needs some way to analyze the computer-readable data and determine what those words, as they were used, meant. This second step typically employs some form of artificial intelligence, and there are several basic approaches researchers have taken to develop a system that can extract meaning from words.

[0004] One such approach involves statistical computational linguistics. This approach relies on the relatively predictable nature of human speech. Statistical computational linguistics begins with a corpus, which is a list of sample utterances contained in the grammar. This corpus is analyzed and statistical properties of the grammar are extracted. These statistical properties are implemented in rules, which are then applied to new, spoken utterances in an attempt to statistically "guess" the meaning of what was said.

[0005] Because of the large number of possible utterances in any language (English, Chinese, German, etc.), no corpus-based language system attempts to list the full set of valid utterances in that language. Some systems, however, have attempted to reduce the number of possible utterances by constraining, or restricting the valid ones to those in a predefined grammar. For example, U.S. Patent 5 652 897, issued to Linebarger, et al, assigned to Unisys Corporation, Blue Bell, Pa. teaches a language processor that only understands air traffic control instructions. There, the air traffic controller's sentence was segmented into individual instructions, which were then individually processed to determine their meaning. Unfortunately, this type of processing can quickly consume much computing power when the valid grammar is increased from the relatively limited vocabulary of air traffic controls to, for example, a bank automated teller machine that can handle all sorts of transactions.

[0006] Other natural language systems may allow for a full range of utterances, but this high degree of generality also requires much computing power. What is needed is a language understanding system that can interpret speech in a constrained grammar that does not require the full generality of a natural language system.

[0007] US-A-5 390 279 relates to partitioning speech rules by context for speech recognition. To this purpose US-A-5 390 279 detects spoken words. As shown in Figs 1 and 2 of US-A-5 390 279 digitized sound signals on line 201 are processed by a first computer 108 in a process called features extraction to separate background noise from speech. The features on lines 211 are not words or characters yet. A second process 220 attempts to recognize the words in the sound features by searching "language models" supplied via line 222. The language models are in effect the vocabulary to be searched. Depending on the features on line 211, the language model Generator/Interpreter calls up vocabulary and rules from application programs 1 to N under the control of the second processor 102 and its operating system is shown as block 244.

[0008] According to US-A-5 390 279 there are two methods for generating a language model. The first method simply uses all of the speech rules to construct a one-time static language model. The problem with this approach is that the language model is so large that the recognizer 220 is slowed down by computational problems and is "more error prone".

[0009] Further, US-A-5 390 279 suggests that the language model be constructed on the fly or dynamically. In the end, analysis the recognizer 220 transmits recognized words 221 to an interpreter process 230 for actions. Thus, when combined, processes 210, 220 and 230 produce words similar to commercially available ASR systems.

[0010] JP-A-08-194600 describes a voice terminal equipment which includes a block 204 called a voice recognizer. Block 201 receives the voice output and converts it to a retrievable command stored in buffer 208. Had the command been a request, block 203 converts the request to an answer output using answer table 207 and a synthesizer block.

[0011] WO-A-97/10589 teaches an automatic call routing system and method including a block 15 called a "Speech Recognizer" and a block 20 called an "Interpretation Module." This system and method employ an unconstrained

vocabulary and grammar which is acquired on the fly. The system selects a meaningful response if it can identify a meaningful phrase in its unconstrained vocabulary.

Summary of the Invention

[0012] A general purpose of the present invention is to provide system and method for providing constrained speech understanding capabilities to an interactive voice recognition system.

[0013] Another object of the present invention is to provide system and method for simplifying the task of interpreting the meaning behind a spoken utterance.

[0014] A further object of the present invention is to provide system and method for creating a corpus-based speech recognition system that is highly accurate in its interpretation of the meaning behind a spoken utterance.

[0015] A still further object of the present invention is to provide a system and method for employing a plurality of runtime interpreters that are connected to the interactive voice response system by a computer network.

[0016] These and other objects are accomplished by the present invention which provides a system and method, as defined in claims 1 and 10. A runtime interpreter receives, as input, an annotated corpus which is a list of valid utterances, context identifiers for each valid utterance, and token data for each valid utterance representing the meaning behind the utterance. The runtime interpreter also receives, as input, an utterance in text form which is to be found in the corpus.

[0017] When the runtime interpreter is given an utterance to interpret, the runtime interpreter searches through the corpus, locates the valid utterance being searched for, and returns the token which represents the meaning of the valid utterance.

[0018] The runtime interpreter also supports the use of variables to reduce the size of the corpus. Some utterances may include numbers, dates, times or other elements that have too many combinations to enumerate in the corpus. For example, the utterance "My birthday is xxx", where 'xxx' is the day of the year, could result in 366 corpus entries, one for each possible day of the year (including leap day). In the present invention, however, a variable would be used to represent the date. Thus, a reduced corpus would include just one entry for this utterance: "My birthday is [DATE]". The runtime interpreter is able to identify these variables in the corpus, and performs additional processing during runtime to interpret the variables. The variable values, once interpreted, are then stored in a predefined data structure associated with the token whose utterance included the variable. This variable value can then be retrieved by the interactive voice response system.

[0019] The present invention also provides a custom processor interface which allows the developer of the interactive voice response system the ability to customize the operation of the runtime interpreter without actually modifying the interpreter itself.

[0020] Furthermore, the present invention provides for a system for using a plurality of interpreters that are connected to a computer network. Distributed interpreters are provided which include the same custom processor interface and runtime interpreter mentioned above. The distributed interpreters, however, include an additional manager for controlling messaging between the distributed interpreter and the computer network. A resource manager is also provided, which keeps track of the distributed interpreters that are connected to the network and manages their use by an interactive voice response system.

Brief Description of the Drawings

[0021]

Figure 1 depicts an overview of an embedded natural language understanding system.

Figure 2 is a table showing the variable types supported in the preferred embodiment.

Figure 3 depicts sample formats for the annotated ASR corpus files and vendor-specific ASP grammar file.

Figure 4 is a flow diagram depicting the operation of the IVR as it accesses the runtime interpreter.

Figure 5 depicts the distributed system architecture.

Description of the Preferred Embodiment

[0022] Before describing the present invention, several terms need to be defined. These terms, and their definitions, include:

annotated ASR corpus file - data file containing a listing of valid utterances in a grammar, as well as token data for each valid utterance which represents the meaning of the valid utterance to the interactive voice recognition system (IVR 130).

automatic speech recognition (ASR) - generic term for computer hardware and software that are capable of identifying spoken words and reporting them in a computer-readable format, such as text (characters).

cells - discrete elements within the table (the table is made up of rows and columns of cells). In the example rule given with the definition of 'rules' below, each of "I want", "I need" and "food" would be placed in a cell. Furthermore, in the preferred embodiment, the cells containing "I want" and "I need" are vertically adjacent to one another (same column). Vertically adjacent cells are generally OR'd together. The cell containing "food", however, would occur in the column to the right of the "I want" and "I need" column, indicating the fact that "food" must follow either "I want" or "I need" and as such, the cell containing "food" will be AND'd to follow the cells containing "I want" and "I need".

constrained grammar - a grammar that does not include each and every possible statement in the speaker's language; limits the range of acceptable statements.

corpus - a large list.

grammar - the entire language that is to be understood. Grammars can be expressed using a set of rules, or by listing each and every statement that is allowed within the grammar.

grammar development toolkit (104) - software used to create a grammar and the set of rules representing the grammar.

natural language understanding - identifying the meaning behind spoken statements that are spoken in a normal manner.

phrase - the "building blocks" of the grammar, a phrase is a word, group of words, or variable that occupies an entire cell within the table.

rules - these define the logic of the grammar. An example rule is: ("I want" | "I need")("food"), which defines a grammar that consists solely of statements that begin with "I want" OR "I need", AND are immediately followed with "food".

runtime interpreter (124) - software that searches through the annotated corpus (122) whenever a valid utterance is heard, and returns a token representing the meaning of the valid utterance.

runtime interpreter application program interface (RIAPI) - set of software functions that serve as the interface through which the interactive voice response system (130) uses the runtime interpreter.

speech recognizer (116) - combination of hardware and software that is capable of detecting and identifying spoken words.

speech recognizer compiler (114) - software included with a speech recognizer (116) that accepts, as input, a vendor-specific ASR grammar file (112) and processes the file (112) for use in a speech recognizer (116) during runtime.

table - two dimensional grid used to represent a grammar. Contents of a table are read, in the preferred embodiment, from left to right.

token - each valid utterance in the table is followed by a cell that contains a token, where the token is a unique data value (created by the developer when s/he develops the grammar) that will represent the meaning of that valid utterance to the interactive voice response system (130).

utterance - a statement.

utterance, spoken - an utterance that was said aloud. The spoken utterance might also be a valid utterance, if the spoken utterance follows the rules of the grammar.

utterance, valid - an utterance that is found within the grammar. A valid utterance follows the rules which define the grammar.

variable - "place holder" used in the corpus (122) to represent a phrase which has too many possibilities to fully enumerate. For example, the utterance "My favorite number between one and a million is xxx" could result in 999,998 corpus entries, one for each possible number. In the present invention, however, a variable would be used to represent the number in the corpus (122). Thus, a reduced corpus (122) would include just one entry for this utterance: "My favorite number between one and a million is [INTEGER]". The runtime interpreter (124) is able to identify this variable in the corpus, and performs additional processing during runtime to interpret the number.

vendor-specific ASR grammar file (112) - a data file that contains the set of rules representing a grammar, and is written in a format that will be recognized by the speech recognizer compiler (114).

[0023] Referring now to the drawings, where elements that appear in several drawings are given the same element number throughout the drawings, the Structures necessary to implement a preferred embodiment of an embedded natural language understanding system (100) are shown in figure 1. The basic elements comprising:

an interactive voice response system (130), or IVR;

the grammar development toolkit (104);

a compiler (114) and speech recognizer (116) that are part of an automatic speech recognition (ASR) system (118);

an annotated automatic speech recognition (ASR) corpus file (122) ;

a vendor-specific ASR grammar file (112);
 the runtime interpreter (124); and
 the custom processor interface (126), or CP; and
 the runtime interpreter application program interface (128), or RIAP. These elements will be discussed in detail
 further below, but an initial overview of the embedded architecture will be helpful to a full understanding of the
 elements and their roles.

1. Overview of Embedded Architecture

[0024] The following overview discusses the embedded architecture, which employs a single runtime interpreter (124). There is a second, distributed, architecture which employs a plurality of runtime interpreters. The distributed architecture will be discussed further below.

[0025] The first step in implementing a natural language system is creating the set of rules that govern the valid utterances in the grammar. As an example, a grammar for the reply to the question: "what do you want for lunch?" might be represented as:

<reply>: ("I want"|"I'd like")("hotdogs"|"hamburgers");

Under this set of rules, all valid replies consists of two parts: 1) either "I want" or "I'd like", followed by 2) either "hot dogs" or "hamburgers". This notation is referred to as Backus-Naur-Form (BNF), a form of grammar that uses logical ANDs and ORs. The preferred embodiment of the present invention generates this type of grammar.

[0026] Referring to Figure 1, the grammar is generated by a developer using the grammar development toolkit (104). In the preferred embodiment, the toolkit (104) is developed using a computer that has an Intel-based central processing unit (CPU 102) such as the Intel Pentium® with Microsoft Visual Basic® as the software development program. The computer also contains random access memory (RAM 106), memory files (108) stored in system memory, and keyboard (110).

[0027] The toolkit (104) is a novel spreadsheet-oriented software package that provides the developer of a natural language application with a simplified way of generating a grammar.

[0028] When the developer has completed the grammar using the toolkit (104), two outputs are generated by the toolkit (104) for use in the natural language system. The first such output is a vendor-specific ASR grammar file (112), which is saved in a format that will be recognizable by the automatic speech recognition system, or ASR (118). The ASR system (118) includes two parts, a compiler (114) and the actual speech recognizer (116). In the preferred embodiment, speech recognizer (116) is a continuous speech, speaker independent speech recognizer. Commercially available speech recognizers (116) include the ASR-1500, manufactured by Lernout & Hauspie; Watson 2.0, manufactured by AT&T; and Nuance 5.0, by Nuance. The preferred embodiment of the toolkit (104) is able to generate grammar files for any of these recognizers.

[0029] The vendor-specific ASR grammar file (112) contains information regarding the words and phrases that the speech recognizer (116) will be required to recognize, written in a form that is compatible with the recognizer. The file is also optimized to take advantage of peculiarities relating to the chosen speech recognizer (116). For example, experience with the L&H recognizers has shown that L&H grammars perform well if the grammar avoids having multiple rules with the same beginning (three rules starting with "I want"). Optimization of a grammar for an L&H recognizer would rewrite a set of rules from <rule1>:(ab)l(ac)l(ad), to <rule2>:a(blcd). Here the three rules of 'rule1' have been rewritten and combined into the one rule of 'rule2'.

[0030] In order to operate and recognize speech, the speech recognizer will need to compile the vendor-specific ASR grammar file (112) using a compiler tool (114) supplied by the ASR system (118) vendor. The preferred embodiment of the toolkit (104) knows, when the grammar is first generated, which speech recognizer (116) will be used and is able to format the vendor-specific ASR grammar file (112) accordingly.

[0031] The second output from the toolkit (104) is an annotated ASP. corpus (122), which is actually a pair of flat files. A sample format for the files is shown in figure 3. The first of the pair is a corpus file, and contains a listing of all possible logical sentences or phrases in the grammar (with the exception of variables, discussed below), the compartments (groups of tables) in which they appear, and a value representing the class of the utterance (sentence) heard. The second is an answers file that maps each utterance class with a token, or data value that represents the meaning of the utterance. These two files will be used by the runtime interpreter (124).

[0032] During runtime, a speaker speaks into the microphone (or telephone) (120) attached to the speech recognizer (116). The recognizer (116) identifies the words and phrases it hears and notifies the IVR (130) when a valid utterance has been heard. The IVR (130) is the system which needs the speech understanding capabilities, and includes the necessary external connections and hardware to function (for example, a banking IVR - 130 might include a connection to the bank database, a keypad for entering data, a visual display for displaying information, a dispenser for dispensing money, and a speaker for speaking back to the user). This valid utterance is passed, in a computer-readable form such as text, to the IVR (130) which then notifies the runtime interpreter (124) of the utterance that was heard. The runtime

interpreter (124) consults the annotated ASR corpus (122) and returns an appropriate token to the IVR (130) for the valid sentence heard by the recognizer (116). This token represents the meaning of the utterance that was heard by the recognizer (116), and the IVR (130) is then able to properly respond to the utterance. The CP (126) and RIAPI (128) serve as software interfaces through which the IVR (130) may access the runtime interpreter (124). It is the IVR (130) that ultimately uses the speech capabilities to interact with the speaker during runtime.

3. The Runtime Interpreter

[0033] The runtime interpreter (124) is a software component that receives, in text form, a valid spoken utterance chat was heard and context information identifying the compartment(s) to be searched. The runtime interpreter (124) then performs a search through the corpus file (122) (which has been loaded into RAM for faster searching) to find the valid utterance. Once a valid utterance is found in the corpus, the associated token is stored in memory to be retrieved by the IVR (130). In an embedded application, calls made to the runtime interpreter (124) are made by functions within the Custom Processor (126), or CP. The CP (126) is another software component that is originally a transparent "middleman" between the runtime interpreter (124) and the RIAPI (128). The IVR (130), created by the developer, only accesses the functions within the RIAPI (128). The RIAPI (128) will make the necessary CP (126) calls which, in turn, make the necessary runtime interpreter (124) calls.

[0034] The purpose for having the CP (126) lies in customizability. The CP (126) can be customized by the developer to enhance the processing of utterances. For example, the developer may wish to perform some type of processing on the utterance before it is actually processed by the runtime interpreter (124). This pre-processing can be added, by the developer, to the CP (126) without actually modifying the runtime interpreter (124). Use of the CP (126) is particularly convenient when the underlying IVR (130) is done in a low level scripting language, such as Vos (by Parity) or BlaBla (by MediaSoft), that does not directly support the pre-processing of the utterance text. If the IVR (130) is written in a higher level language, such as C++, then pre-processing of the utterance text can be done in the IVR (130) code itself, without need for the CP (126).

[0035] The runtime interpreter (124) also provides functionality to extract variables from utterances. When the corpus file is first loaded, corpus items that contain variables are flagged. If an initial binary search through the corpus fails to find the exact utterance, a second search is performed to find a partial match of the utterance. This time, only flagged corpus items are searched, and a partial match is found if the utterance contains at least the non-variable portions of a corpus item.

[0036] For example, the preferred embodiment corpus file (122) format uses square brackets '[' and ']' to set off variables from normal words in the valid utterances. Thus, the utterance "I want to transfer [CURRENCY1, money] to savings" might be found in the corpus file. If the spoken utterance heard by the recognizer (116) is "I want to transfer ten dollars to savings", an initial binary search would probably fail to match the spoken utterance with any of the corpus items. If this initial search fails, the interpreter (124) then performs a second search of all flagged corpus items. The spoken utterance that was heard at least contained "I want to transfer ... to savings", and a partial match would be made. The unmatched words, "ten dollars", would then be processed by another algorithm as a variable of type [CURRENCY1, money] which would convert the phrase "ten dollars" to 10.00 and return 10.00 as the variable associated with the token "money". This variable data is then stored in a predefined data structure that is associated with the location in memory where the token was stored. When the IVR (130) processes the token, it knows that variable data was also returned and retrieves the variable data from memory.

[0037] The algorithm for converting variables in utterances to variable data depends on the type of data contained within the variable. Figure 2 shows variable types that are supported by the preferred embodiment. The following pseudocode illustrates the steps used in the preferred embodiment to convert the variable portion of the utterance (in text form) to variable data (in number form).

INTEGER1: ("One hundred thousand and ten")

Set TEMP result buffer = 0;

5 Separate variable portion of utterance into individual
words

(i.e. "one" "hundred" "thousand" "and" "ten") based

10 on blank spaces between words.

FOR each individual word (reading left to right):

15

20

25

30

35

40

45

50

55

```

    if (individual word = "one"), increase TEMP by 1;
    if (individual word = "two"), increase TEMP by 2;
5      ...
    if (individual word = "twenty"), increase TEMP by
20;
10    if (individual word = "thirty"), increase TEMP by
30;
    ...
15    if (individual word = "ninety"), increase TEMP by
90;
    if (individual word = "hundred")
20      if (TEMP > 1000) ("four thousand five hundred",
when the
        word "hundred" is reached, will have been
25 handled as
        "four thousand five", and TEMP would be 4005.
This
30      needs to be changed to 45, before multiplying
by 100)
        TEMP = (TEMP/100) + least significant digit of
35 TEMP;
        end if;
        multiply TEMP by 100;
40    end if;
    if (individual word = "thousand"), multiply TEMP by
1000;
45    if (individual word = "and"), ignore;
    END FOR loop;

```

50 INTEGER2: ("One Two Three Four")

As in INTEGER1, break up variable utterance into
individual


```

        words, and set a TEMP buffer = 0;
    FOR each individual word (reading left to right)
5      multiply TEMP by 10;
        if (individual word = "one"), increase TEMP by 1;
        ...
10     if (individual word = "nine"), increase TEMP by 9;
    END FOR;

```

```

15  CURRENCY1: ("Twenty three dollars and fifteen cents")
    As in INTEGER1, break up variable utterance into
    individual

```

```

20      words, and set a TEMP buffer = 0;
    FOR each individual word (reading left to right):
        if (individual word = "one"), increase TEMP by 1;
25      if (individual word = "two"), increase TEMP by 2;
        ...
        if (individual word = "twenty"), increase TEMP by
30      20;
        if (individual word = "thirty"), increase TEMP by
30      30;
        ...
35      if (individual word = "ninety"), increase TEMP by
        90;
40      if (individual word = "hundred")
            if (TEMP > 1000)
                TEMP = (TEMP/100) + least significant digit of
45      TEMP;
            end if;
            multiply TEMP by 100;
50      end if;
        if (individual word = "thousand"), multiply TEMP by
        1000;

```

```

    if (individual word = "and"), ignore
    if (individual word = "dollars"), multiply TEMP by
5 100;
    if (individual word = "cents"), ignore
    END FOR
10 Return the number TEMP/100;

```

```

15 CURRENCY2: ("two three dollars and one five cents")
    As in INTEGER1, break up variable utterance into
    individual
    words, and set a TEMP buffer = 0;
20 FOR each individual word (reading left to right)
    if (individual word = "dollars"), multiply TEMP by
25 100;
    else, multiply TEMP by 10;
    if (individual word = "one"), increase TEMP by 1;
    ...
30 if (individual word = "nine"), increase TEMP by 9;
    if (individual word = "cents"), divide TEMP by 10;
    if (individual word = "and"), divide TEMP by 10;
35 END FOR;
    Return TEMP/100 as number in dollars;

```

```

40 TIME: ("one o'clock p.m.", "thirteen hundred hours")
    As in INTEGER1, break up variable utterance into
    individual
45 words, and set buffers HOUR, MIN = 0)
    Discard ("o'clock", "minutes", "hours")
    if (first word = "one"), HOUR = 1;
50 ...
    if (first word = "twenty")
        HOUR = 20;
55

```

```

    if (second word = "one"), increase HOUR by 1;
    if (second word = "two"), increase HOUR by 2;
5    if (second word = "three"), increase HOUR by 3;
    next word = third word;
else, next word = second word;
10  FOR (each individual word from next word on)
    if (individual word = "one"), increase MIN by 1;
    ...
15    if (individual word = "twenty"), increase MIN by 20;
    if (individual word = "thirty"), increase MIN by 30;
    if (individual word = "forty"), increase MIN by 40;
20    if (individual word = "fifty"), increase MIN by 50;
    if ((individual word = "p.m.") and (HOUR < 12))
        increase HOUR by 12;
25    end if;
end FOR
if (TIME2), return "HOUR:MIN";
30 if (TIME1)
    if (HOUR > 12)
        decrease HOUR by 12;
35    return "HOUR:MIN PM";
    else
        return "HOUR:MIN AM";
40    end if;
end if;

```

```

45  DATE1: ("March first nineteen ninety seven", "the first
of March")

```

```

    As in INTEGER1, break up variable utterance into
50 individual
        words, and set buffers MONTH, DAY, YEAR, UNKNOWN = 0
and set

```

```

    flag DONE = N)
FOR each word
5   if (word = "january"), MONTH = 1;
    ...
10  if (word = "december"), MONTH = 12;
    if (word = "first", . . . "twentieth", or
"thirtieth")
15      if (word = "first"), DAY = 1;
        ...
        if (word = "twentieth"), DAY = 20;
        if (word = "thirtieth"), DAY = 30;
20      if (UNKNOWN is not 0) (i.e., was "twenty" or
"thirty")
        Add UNKNOWN to DAY;
25      reset UNKNOWN to 0;
        end if;
    else if (word = "oh", "one", . . . "nineteen")
30      if (word = "oh") AND (UNKNOWN is not 0)
        Add (UNKNOWN*100) to YEAR;
        UNKNOWN = 0;
35      go to next word;
    else
        if (YEAR is not 0)
40      Add (value of word) to YEAR;
        go to next word;
        else
45      if (UNKNOWN is not 0) AND ((value of
word)<10)
        Add (value of word) and UNKNOWN to YEAR;
50      UNKNOWN = 0;
        else if (UNKNOWN is not 0)

```

Add (100*UNKNOWN) and (value of word) to
YEAR;

5

 UNKNOWN = 0;

 else

 UNKNOWN = (value of word);

10

 go to next word;

 end else;

 end else;

15

 end else;

 else if (word = "twenty" or "thirty")

 if (UNKNOWN is not 0)

20

 Add (100*UNKNOWN) and (value of word) to YEAR;

 UNKNOWN = 0;

 go to next word;

25

 else

 UNKNOWN = (value of word);

 go to next word;

30

 end else;

 else if (word = "forty", "fifty", . . . "ninety")

 if (UNKNOWN is not 0)

35

 YEAR = 100*UNKNOWN;

 UNKNOWN = 0;

 end if;

40

 if (YEAR is not 0)

 Add (value of word) to YEAR;

 go to next word;

45

 else

 UNKNOWN = (value of word);

 go to next word;

50

 end else;

 else if (word = "hundred")

 if (UNKNOWN is not 0)

55

```

        Add (100*UNKNOWN) to YEAR;
5      UNKNOWN = 0;
        end if;
        go to next word;
10     else if (word = "thousand")
        if (UNKNOWN is not 0)
            Add (1000*UNKNOWN) to YEAR;
15     UNKNOWN =0;
        end if;
        go to next word;
20     end else if;
        if (UNKNOWN is not 0), Add UNKNOWN to YEAR;
        Return MONTH, DAY and YEAR in whatever format is
25     selected;
        end FOR;
    END DATE;
30

```

[0038] The basic operation of the runtime interpreter (124), as seen and used by an interactive voice response (IVR) system, is shown in figure 4. In the following description, the specific function names used are those found in the preferred embodiment. First, the IVR (130) must be started in step 400. The IVR (130) is a software system for performing other duties, such as controlling a bank's automated teller machine, that takes advantage of the speech understanding capabilities of the present invention. For example, a bank may develop an IVR (130) to provide for a talking automated teller machine. In the preferred embodiment, the IVR (130) is responsible for managing the speech recognizer (116) during runtime.

[0039] One of the first things the IVR (130) will want to do is initialize the speech recognizer in step 402. The exact steps necessary for initializing a speech recognizer will vary depending on which commercial speech recognizer is used, but the general steps will involve compiling the vendor-specific ASR grammar (112) that was generated using the toolkit (104), and loading the compiled version into some form of local memory accessible to the speech recognizer (116).

[0040] Next, in step 404, the runtime interpreter (124) will need to be initialized. This is done when the IVR (130) calls the NL_Init function. This function essentially receives, as input, a file path and name for the annotated ASR corpus (122) that will be used for the current application and stores this file path and name in memory.

[0041] In step 406, the IVR (130) finishes setting up the runtime interpreter (124) by calling the NL_OpenApp function. This function access the corpus file whose name and file path were stored by the NL_Init function in step 404, and loads the corpus into system memory (RAM) in preparation of being searched. In order to optimize the search, the contents of the corpus file (the various valid utterances) are alphabetized when loaded into RAM. Alphabetizing the valid utterances will enhance the search performance because, in the preferred embodiment, a binary search is used to match an utterance with a token. Binary searches are a common method of searching through sorted lists to find a target element, and basically involves progressively halving the range of list items being searched until the target item is found.

[0042] During this loading process, the corpus data is also optimized by 1) flagging corpus items that contain variables and 2) generating the list (from large to small) that specifies the order in which corpus items are processed for the second search. This last bit of optimization is important because, as the second search looks for fragments, smaller fragments (fewer words) may inadvertently match when a larger fragment is more appropriate. For example, the item:

"I want to transfer ... to savings" is smaller than the item "I want to transfer ... British pounds to savings". If the spoken utterance is "I want to transfer ten British pounds to savings" and the smaller item is processed first, it will incorrectly match ("I want to transfer ... to savings" is found) and send the remaining words ("ten British pounds") for processing as a variable in the first item, when "ten" should actually be processed as a variable in the second item. It is important that larger items are processed first when the second search is conducted, and this ordering is done when the corpus is initially loaded into the RAM memory. A separate list of pointers is generated and stored in memory when the corpus is loaded, and this list identifies the order (large to small) in which items with variables should be processed. A list of flagged corpus items is also stored in memory.

[0043] Once both the speech recognizer (116) and runtime interpreter (124) have been initialized, and after the runtime interpreter (124) has loaded the corpus, the runtime interpreter is ready to do its job. At this point, the IVR (130) may have other processing to do, and the runtime interpreter (124) waits.

[0044] At some point in the future, the IVR (130) will detect that a conversation with a speaker is about to begin. When this happens, the IVR (130) will need to open a session within the runtime interpreter (124) (a session is a dialog exchange with the speaker). The IVP (130) does this by calling the NL_OpenSession function in step 406. This function creates a session handle and associates the session handle with the session that was opened. Future function calls relating to this session will use the session handle to reference the session.

[0045] Then, in step 408, the speech recognizer (116) informs the IVR (130) that a complete utterance may have been heard. In the preferred embodiment, speech recognizers (116) are of the type that return data-in NBest form. NBest form is simply an output data format that includes a list of possible valid utterances (in text form) heard by the speech recognizer (116) along with a confidence number indicating the likelihood that each valid utterance was heard.

[0046] The NBest format is helpful when there are multiple valid utterances that sound alike. For example, if the valid grammar includes "I want honey" and "I want money", and the speaker mumbles "I want mfhoney", the speech recognizer will return both valid utterances as possibilities, rather than simply returning the single valid utterance that it believed sounded most correct. A confidence number is also included for each valid utterance, indicating the speech recognizer's confidence that that particular valid utterance was indeed the one it heard. This plurality of possibilities is helpful when the runtime interpreter (124) also knows the context of the current discussion and can use the context information to more accurately determine which valid utterance was meant. As will be described below, the preferred embodiment of the runtime interpreter (124) will use such context information in its determination of what was meant.

[0047] After the IVR (130) receives the output from the speech recognizer (116), the output is then passed, in step 410, to the runtime interpreter (124) for interpretation. To do so, the IVR (130) will call, in the preferred embodiment, the NL_AnalyzeNBest function. This function accepts as input the NBest data received by the IVR (130), a session handle, and a context pointer indicating the compartment that is to be searched.

[0048] When the NL_AnalyzeNBest function is executed, the runtime interpreter (124) then searches through the corpus (122) that has been loaded into memory to find the valid utterance. If a match is found, the return token is stored in memory. If no match is found, the variables search discussed above will be performed and the variable data will be stored in a predefined data structure. This search is shown in step 412.

[0049] After calling NL_AnalyzeNBest, the IVR (130) will need to call NL_GetResult, in step 416, to retrieve from memory the token stored by the NL_AnalyzeNBest function. If the token indicates that a variable was included in the utterance, then the IVR (130), in step 416, will call NL_GetVariable to retrieve the variable values from the predefined data structure in memory used by NL_AnalyzeNBest to store the variable data.

[0050] Once the token and any necessary data have been stored in memory, the runtime interpreter (124) is finished for the session (for now). In step 418, the runtime interpreter (124) waits for either another utterance or an end to the session.

[0051] If another utterance occurs, the speech recognizer (116) will again notify the IVR (130) in step 408, the IVR (130) will call NL_AnalyzeNBest in step 410, and the process continues as it did before.

[0052] If the session is to end, the IVR (130) will call NL_CloseSession in step 420. Closing the session deassociates the session handle.

[0053] At this point, step 422, the runtime interpreter (124) waits for either a new session to begin or for the command to shut down the current application. If a new session is to begin, the IVR (130) will call NL_OpenSession again in step 404 and processing continues from step 404 as before. If the current application is to be shut down, then IVR (130) will call NL_CloseApp in step 424 to release the memory that had been allocated when the application was opened.

[0054] Then, in step 426, the IVR (130) calls NL_Shutdown to undo the effects of NL_Init.

[0055] Finally, in steps 428 and 430, the IVR (130) is responsible for shutting down the speech recognizer (116) as well as the IVR (130) itself. The actual steps necessary will vary depending on the selected speech recognizer (116) as well as the IVR developer.

[0056] The runtime interpreter (124) also provides functionality for the developer who wishes to manage the Nbest data passed by the CP (126). Functions are available to create Nbest buffers (NB_CreateBuffer); create an Nbest

buffer with only one utterance (NB_CreateOneBest); set an utterance in an Nbest buffer (NB_SetUtterance); set a score for an utterance in an Nbest buffer (NB_SetScore); set an utterance/score pair in an Nbest buffer (NB_SetUtteranceScore); determine the number of utterances that can be stored in the Nbest buffer (NB_GetNumResponses); get an utterance from an Nbest buffer (NB_GetUtterance); get a score from an Nbest buffer (NB_GetScore) and release the memory allocated for a specified Nbest buffer (NB_DestroyBuffer).

4. The Runtime Interpreter Application Program Interface

[0057] The runtime interpreter application program interface (128), or RIAPI, is the set of software functions actually used by the developer of the IVR (130) to interact with the runtime interpreter (124). The functions which are included in the preferred embodiment of the RIAPI (128) include: NL_Init(), NL_OpenApp(), NL_OpenSession(), NL_AnalyzeNbest(), NL_GetResult(), NL_GetVariable(), NL_CloseSession(), NL_CloseApp() and NL_Shutdown().

[0058] NL_Init is an initialization function that is called one time during startup to process initialization information and allocate memory for sessions. Initialization information can include a name for a local log file, the maximum number of sessions and the routing mode (embedded or distributed - distributed architecture will be discussed further below). A call to NL_Init, in the exemplary embodiment, results in a call to CP_Init (the CP equivalent), which then calls SAI_Init (the runtime interpreter 124 equivalent). Most of the following RIAPI (128) functions will also result in function calls to the CP (126), which then calls the corresponding runtime interpreter (124) function. Two exceptions in the preferred embodiment are the NL_GetVariable and NL_GetResult function, which directly access memory to retrieve the variable or result.

[0059] NL_OpenApp is called to establish an application in the interpreter (124). As stated before, an application is an instance, or Implementation, of a project. Opening an application causes the interpreter (124) to load the corpus files (122) associated with the application.

[0060] NL_OpenSession is called when a session is desired under an open application. A session is essentially a conversation with a speaker, and it is possible for several sessions to exist for the same application (if the IVR 130 manages several speech recognizers, for example).

[0061] NL_AnalyzeNbest is called by the IVR (130) when the speech recognizer has indicated that it has Nbest output ready. The IVR (130) calls this function to send this Nbest output, as well as contextual information, to the runtime interpreter (124) for analysis.

[0062] NL_GetResult is called by the IVR (130) to read the token which was stored in memory by the runtime interpreter (124).

[0063] NL_GetVariable is called when the token stored by the interpreter (124) is of a type that has variable data associated with it. The call to NL_GetVariable retrieves this variable data from a memory data structure used by the interpreter (124) to store the data.

[0064] NL_CloseSession is called to close the specified session and return any allocated resources that were associated with the session. Calling this function may result in the calling of other functions that are also necessary for closing the session. For example, in the embedded architecture, NL_CloseSession calls CP_CloseSession to allow the CP (126) and runtime interpreter (124) an opportunity to properly close their respective sessions and return allocated resources that they no longer need.

[0065] NL_CloseApp is called to close the specified application. This function checks to ensure that all sessions have been closed, and may also call other functions such as CP_CloseApp to allow the CP (126) and interpreter (124) the opportunity to "clean up after themselves" as well.

[0066] NL_Shutdown is called to essentially return the system to the state that existed before NL_Init was called. CP_Shutdown may also be called to have the CP (126) and interpreter (124) deallocate their resources.

[0067] In addition to these basic functions, the RIAPI (128) is also provided with inter/intranet capabilities. If the natural language system is connected via TCP/IP to a network, the TcpCallback function can be used to process asynchronous TCP/IP socket events. The following RIAPI calls designed to support connections through Server Interface Process (SIP) to the internet are also available (although not necessary for non-SIP systems):

NL_WEBConnect (to open a session with a remote web browser user), NL_ReportWEBText (to pass text responses to the interpreter 124), NL_WEBPlay (to present or display file contents to the remote user), NL_WEBListen (to direct one session to accept input from the SIP instance connected by NL_WEBConnect), NL_GetWEBResult (to retrieve results of an NL_WEBListen call) and NL_CloseWEBSession (to close a session).

[0068] As an interface between the IVR (130) and (ultimately) the runtime interpreter (124), the specific calls made to the RIAPI (128) will be dictated by the needs of the IVR (130) for the functionality of the runtime interpreter (124).

5. Overview of the Distributed Architecture

[0069] Thus far, this specification has been describing the elements of an embedded system architecture. In an embedded architecture, both the runtime interpreter (124) and RI-API (128) are software elements that reside on the same computer.

[0070] In a distributed architecture, a plurality of distributed runtime interpreters (508) is located among a plurality of locations within a computer network (in the preferred embodiment, both Unix and Windows NT networks are supported). By having this plurality of interpreters (508), the IVR (130) is able to have a number of utterances processed simultaneously. The clearest advantage to this is the ability to operate multiple sessions at the same time.

[0071] Figure 5 shows the elements of a distributed system architecture. Most of the elements are the same as the ones found in the embedded architecture. Both the grammar (112) and corpus (122) are the same as those used in the embedded architecture. The differences are the plurality of distributed interpreters (508), the resource manager (510 - RM), logger (512), operator display (514) and log viewer (516). The distributed interpreters (508) and RM (506) are discussed further below.

[0072] The logger (512) is simply a software device that records the various messages that are sent between the resource manager (510) and various interpreters (508). Operator display (514) and log viewer (516) are means by which the developer may monitor the operation of the IVR (130) and the various interpreters connected to the system. In the preferred embodiment, the logger (512), operator display (514) and log viewer (516) do not allow the user or operator any control over the IVR (130) application. These devices merely provide information on the operation of the application.

6. The Distributed Interpreter

[0073] In an alternative embodiment of the present invention, a distributed system is used. The distributed system operates on a networked computer system. A networked computer system simply means a plurality of computers, or nodes, which are interconnected to one another via a communications network.

[0074] In a distributed system, each node that performs interpreting duties has a DI manager (504), a DICP (506; and a DI runtime interpreter (508). The DICP (506) and DI runtime interpreter (508) have the same functionality as the CP (126) and runtime interpreter (124) found in the embedded architecture discussed above. The DI manager (504) is another piece of software that is responsible for message processing and coordination of the interpreting duties of the node. Message processing depends on the type of network used to connect the node to the resource manager (510). However, the same general message types are used. The message types and purposes are discussed below.

[0075] The manager (504) itself is a software component, and before it can process any messages it must first be executing on the interpreting node. When the manager (504) is started, it will look in an initialization file for information regarding the application supported by the manager (504). This information includes the name of the application supported, and the file path to the location of the annotated corpus (122) to be used for the application supported.

[0076] The <initialize> message causes the DI manager (504) to initialize the DICP (506) by calling CP_Init, and the DICP (506) initializes the DI runtime interpreter (508) by calling SAI_Init. This message also causes the DI manager (504) to initialize the application to be supported, by calling CP_OpenApp and SAI_OpenApp to open the application. As discussed above, opening an application requires loading the corpus (122). The location of the corpus (122) to be loaded is passed on to the DI runtime interpreter (508). When the DI runtime interpreter (508) completes its initialization (and the corpus 122 is loaded), it generates an application handle which is a data object that references the current application. This handle is returned to the DICP (506), which in turn passes it back to the DI manager (504). Whenever an error occurs within the DI (502), the DI manager (504) composes a <tell error> message describing the error and returns it to the RM (510).

[0077] A session will be opened when the DI manager (504) receives a <start session> message. This message includes a resource address which identifies the sending IVR (130) and a session identifier. The DI manager (504) checks to make sure there is not already a session opened with the same resource address, and if there is not, creates a session object which will represent the session. A session object is essentially a handle, similar to the application handle discussed above, that references this session. The DI manager (504) then opens the session in the DICP (506) and DI runtime interpreter (508) by calling the CP_OpenSession function, which calls the SAI_OpenSession function. The return value of SAI_OpenSession is passed back to CP_OpenSession, which returns it to the DI manager (504). Again, errors are reported by the DI manager (504) with a <tell error>. message.

[0078] Once a session has been opened, the DI (502) is ready to interpret. There are two messages which can start the process of interpretation. First, the DI manager (504) could receive an <analyze> message. An <analyze> message contains all the context and nbest information normally needed for CP_AnalyzeNbest. The DI manager (504) then calls the DI runtime interpreter (508) convenience functions NB_CreateBuffer and NB_SetUtteranceScore to prepare a structure with the context and nbest data. The DI manager (504) then provides this data structure as input to the

CP_AnalyzeNbest function, which calls the SAI_AnalyzeNbest function which performs the search described above for the embedded architecture. When these functions have completed, their return values propagate back to the DI manager (504), which composes and sends a <reply> message back to the RM (510).

[0079] Receiving the <analyze> message was just one way the interpretation could be started. The other way occurs when the context and nbest data are sent in separate messages. When this occurs, the RM (510) sends a first message, or <state> message, containing the context and a resource address identifying the session in which the utterance was heard. Upon receipt of this message, the DI manager (504) first confirms that the resource address is indeed that of an existing session. If it is, the DI manager (504) retrieves the session handle associated with the resource address, and stores the context information from the message in a temporary memory area to await further processing.

[0080] This further processing will occur when the second message is received by the DI manager (504). The second message, or <nbest> message, contains a resource address and some nbest data. When the <nbest> message is received, the DI manager (504) again checks to make sure the resource address included in the <nbest> message is that of an existing session. If so, the DI manager (504) then looks to the temporary memory area associated with the session, and finds the previously stored context information. Taking the nbest data and context data, the DI manager (504) then makes a call to CP_AnalyzeNbest, which then calls SAI_AnalyzeNbest, where the corpus (122) is searched to find the token associated with the utterance in the nbest data.

[0081] A session is ended when the DI manager (504) receives the <lost call> message. This message includes a resource address, and the DI manager (504) checks to make sure that the resource address does indeed reference an open session. If so, the DI manager (504) calls CP_CloseSession, which then calls SAI_CloseSession, and the session is closed much in the same way a session is closed in the embedded architecture.

[0082] If the entire application, is to be shut down, the DI manager (504) will receive a <terminate> message. Since each manager (504) can only support one application at a time, shutting down an application is the same as shutting down the manager (504). When the DI manager (504) receives this message, it makes the necessary calls to CP_CloseSession to close any remaining sessions that are open, and finally calls CP_Shutdown, which calls SAI_ShutDown, and all resources allocated to the manager (504), DICP (506) and DI runtime interpreter (508) are released.

7. The Resource Manager

[0083] The resource manager (510) monitors the operation of the various distributed interpreters (508) connected to the network, and distributes RIAPI (128) requests among the interpreters (508). In the preferred embodiment, the RM (510) receives a message whenever a distributed interpreter (508) is initiated and records the application that is supported by the distributed interpreter (508). Then, as the resource manager receives requests from the IVR(s) (130) through the RIAPI (128), it checks to see which distributed interpreter (508) can handle the request (supports the application) and formulates a message containing the IVR (130) request and sends it to the appropriate manager (504) for processing. The resource manager (510) communicates with managers (504) using the messages described above.

[0084] In light of the above teachings, it is understood that variations are possible without departing from the scope of the invention embodied in these teachings. Any examples provided as part of the inventors' preferred embodiment are presented by way of example only, and are not intended to limit the scope of the invention. Rather, the scope of the invention should be determined using the claims below.

[0085] Where technical features mentioned in any claim are followed by reference signs, those reference signs have been included just for the sole purpose of increasing intelligibility of the claims and accordingly, such reference signs do not have any limiting effect on the scope of each element identified by way of example by such reference signs.

Claims

1. A computer system for identifying the meaning behind a valid spoken utterance in a known grammar, said system comprising:

- a central processing unit (CPU) (102);
- a system memory coupled to said CPU (102) for receiving and storing memory files (108);
- a random access memory (RAM) portion (106) of said system memory, for temporarily receiving and storing data during operation of said CPU (102);
- a predetermined fixed annotated automatic speech recognition (ASR) corpus file (122), stored in said system memory, containing a listing of all expected valid utterances in said known grammar and stored therewith token data representing the meaning of each of said listed expected valid utterances;
- an automatic speech recognition (ASR) system (114, 116, 118) coupled to said CPU (102) for detecting spoken

utterances and for generating digital output signals indicative of detected valid utterances;
 a predetermined fixed vendor specific ASR grammar file (112), stored in said system memory, coupled to said ASR system (114, 116, 118) and containing data representing all valid utterances to be detected by said ASR system (114, 116, 118) in a form that is compatible with the ASR system (114, 116, 118); and
 5 runtime interpreter means (124) coupled to said CPU (102) and said annotated ASR corpus file (122) for identifying, when said CPU (102) receives said ASR system digital output signals that are indicative of a detected valid utterance, the meaning behind said detected valid utterance by searching through the contents of said annotated ASR corpus file (122), and for returning said identified meaning as said token data.

10 2. The system of claim 1, wherein said runtime interpreter means (124) further includes means for performing a comparison search through said annotated ASR corpus file (122) in said system memory to find the token data identifying the means behind said detected valid utterance.

15 3. The system of claim 2, wherein said runtime interpreter means (124) further includes means for performing a partial match search through the contents of said annotated ASR corpus file (122) upon failure of said comparison search, where said partial search seeks a partial match of said detected valid utterance in said annotated ASR corpus file (122).

20 4. The system of claim 3, wherein said runtime interpreter means (124) further includes variable processing means for processing the unmatched portion of said detected valid utterance as a variable to identify the meaning behind said unmatched portion.

25 5. The system of claim 4, wherein said variable processing means generates variable data representing the meaning of said unmatched portion of said detected valid utterance.

6. The system of one or more of claims 1-5, further comprising a runtime interpreter application program interface (RIAPI) means (128), coupled to said runtime interpreter means (124), for accessing said runtime interpreter means (124).

30 7. The system of claim 6, wherein said RIAPI means (128) further comprises a custom processor (CP) interface (126), coupled to said RIAPI means (128) and said runtime interpreter means (124), and an interactive voice response (IVR) system (130) used by said RIAPI means (128) to access said runtime interpreter means (124).

35 8. The system of claim 7, wherein said IVR system (130) is coupled to said ASR system (114, 116, 118) and said runtime interpreter means (124) for searching said annotated ASR corpus file (122) to find said token data representing the meaning behind said detected spoken utterance.

40 9. The system one or more of claims 1-8, where said computer system comprises a network system with a plurality of said runtime interpreter means (508a ... 508n) distributed on a plurality of computers on said computer system network.

10. A method for identifying the meaning behind a valid spoken utterance in a known grammar, comprising the steps of:

45 loading a predetermined fixed annotated automatic speech recognition (ASR) corpus file (122) into a system memory of a computer, where said fixed annotated ASR corpus file (122) contains a listing of all expected valid utterances in said fixed grammar and token data representing the meaning of each of said listed expected valid utterances;

loading a predetermined fixed vendor specific ASR grammar file (112) into said system memory, containing data representing all valid utterances to be detected by an ASR system (114, 116, 118) in a form that is compatible with the ASR system (114, 116, 118);

50 detecting a valid utterance by said automatic speech recognition (ASR) system (114, 116, 118);

initiating a request to search in said fixed annotated ASR corpus file (122) for occurrence of said detected valid utterance;

55 performing a search, by routine interpreter means (124), through the contents of said fixed annotated ASR corpus file (122) loaded in said system memory to find a valid utterance among said listed expected valid utterances corresponding to said detected valid utterance, and identifying the meaning behind said detected valid utterance as the token data associated to the found valid utterance and

returning said token data corresponding to the meaning behind said detected valid utterance in said fixed

annotated ASR corpus file (122) to the requestor.

Patentansprüche

1. Ein Computersystem zur Identifizierung der Bedeutung einer gültigen gesprochenen Äußerung in einer bekannten Grammatik, wobei das System folgendes umfasst:

eine Zentraleinheit (CPU) (102);
 einen Systemspeicher, der mit der CPU (102) verbunden ist, um Speicherdateien (108) zu empfangen und zu speichern;
 einen Direktzugriffsspeicher-(RAM)-Abschnitt (106) des Systemspeichers, um temporär Daten zu empfangen und zu speichern, während des Betriebs der CPU (102);
 eine vorherbestimmte feste kommentierte automatische Spracherkennung-(ASR)-Corpusdatei (122), die im Systemspeicher gespeichert ist, die eine Aufzählung aller erwarteten gültigen Äußerungen in der bekannten Grammatik enthält und damit gespeichert Zeichendaten, die die Bedeutung einer jeden der aufgelisteten erwarteten gültigen Äußerungen darstellen;
 ein automatisches Spracherkennung-(ASR)-System (114, 116, 118), das mit der CPU (102) verbunden ist, um gesprochene Äußerungen zu detektieren, und um digitale Ausgabesignale zu erzeugen, die die detektierten gültigen Äußerungen anzeigen;
 eine vorherbestimmte feste verkäuferspezifische ASR-Grammatikdatei (112), die im Systemspeicher gespeichert ist, die mit dem ASR-System (114, 116, 118) verbunden ist und die Daten enthält, die alle gültigen Äußerungen darstellen, die durch das ASR-System (114, 116, 118) detektiert werden sollen, und zwar in einer Form, die mit dem ASR-System (114, 116, 118) kompatibel ist; und
 ein Laufzeit-Interpreter-Mittel (124), das mit der CPU (102) und mit der kommentierten ASR-Corpusdatei (122) verbunden ist, zur Identifizierung - wenn die CPU (102) die digitalen Ausgabesignale des ASR-Systems empfängt, die eine detektierte gültige Äußerung anzeigen - der Bedeutung der detektierten gültigen Äußerung, indem durch die Inhalte der kommentierten ASR-Corpusdatei (122) gesucht wird, und um die identifizierte Bedeutung als die Zeichendaten zurückzuführen.

2. Das System gemäß Anspruch 1, worin das Laufzeit-Interpreter-Mittel (124) weiterhin ein Mittel einschließt, um eine Vergleichsuche durch die kommentierte ASR-Corpusdatei (122) in Systemspeicher durchzuführen, um die Zeichendaten zu finden, die das Mittel in der detektierten gültigen Äußerung identifiziert.

3. Das System gemäß Anspruch 2, worin das Laufzeit-Interpreter-Mittel (124) weiterhin ein Mittel einschließt, um eine partielle Übereinstimmungssuche durch die Inhalte der kommentierten ASR-Corpusdatei (122) durchzuführen beim Versagen der Vergleichsuche, worin die partielle Suche eine partielle Übereinstimmung der detektierten gültigen Äußerung in der kommentierten ASR-Corpusdatei (122) sucht.

4. Das System gemäß Anspruch 3, worin das Laufzeit-Interpreter-Mittel (124) weiterhin ein variables Verarbeitungsmittel einschließt, um den nicht übereinstimmenden Abschnitt der detektierten gültigen Äußerung als eine Variable zu verarbeiten, um die Bedeutung des nicht übereinstimmenden Abschnitts zu identifizieren.

5. Das System gemäß Anspruch 4, worin das variable Verarbeitungsmittel variable Daten erzeugt, die die Bedeutung des nicht übereinstimmenden Abschnitts der detektierten gültigen Äußerung darstellen.

6. Das System nach einem oder mehreren der Ansprüche 1-5, das weiterhin ein Laufzeit-Interpreter-Anwendungsprogramm-Schnittstellen-(RIAPI)-Mittel (128) umfasst, das mit dem Laufzeit-Interpreter-Mittel (124) verbunden ist, um auf das Laufzeit-Interpreter-Mittel (124) zuzugreifen.

7. Das System gemäß Anspruch 6, worin das RIAPI-Mittel (128) weiterhin eine kundenspezifische Prozessor-(CP)-Schnittstelle (126) umfasst, die mit dem RIAPI-Mittel (128) und dem Laufzeit-Interpreter-Mittel (124) verbunden ist, und ein interaktives auf Sprache reagierendes (IVR)-System (130) umfasst, das durch das RIAPI-Mittel (128) verwendet wird, um auf das Laufzeit-Interpreter-Mittel (124) zuzugreifen.

8. Das System gemäß Anspruch 7, worin das IVR-System (130) mit dem ASR-System (114, 116, 118) und dem Laufzeit-Interpreter-Mittel (124) verbunden ist, um die kommentierte ASR-Corpusdatei (122) zu durchsuchen, um die Zeichendaten zu finden, die die Bedeutung der detektierten gesprochenen Äußerung darstellen.

9. Das System nach einem oder mehreren der Ansprüche 1-8, worin das Computersystem ein Netzwerksystem mit einer Vielzahl von Laufzeit-Interpreter-Mittel (508a ... 508n) umfasst, die auf eine Vielzahl von Computern im Computersystemnetzwerk verteilt sind.

5 10. Ein Verfahren zur Identifizierung der Bedeutung in einer gültigen gesprochenen Äußerung in einer bekannten Grammatik, das die folgenden Schritte umfasst:

das Laden einer vorherbestimmten festen kommentierten automatischen Spracherkennung-(ASR)-Corpusdatei (122) in einen Systemspeicher eines Computers, worin die feste kommentierte ASR-Corpusdatei (122) eine Auflistung aller erwarteten gültigen Äußerungen in der festen Grammatik und Zeichendaten enthält, die die Bedeutung einer jeden der aufgelisteten erwarteten gültigen Äußerungen darstellen;
 10 das Laden einer vorherbestimmten festen verkäuferspezifischen (ASR)-Grammatikdatei (112) in den Systemspeicher, die Daten enthält, die alle gültige Äußerungen darstellen, die durch ein (ASR)-System (114, 116, 118) in einer Form zu detektieren sind, die mit dem ASR-System (114-116, 118) kompatibel ist;
 15 das Detektieren einer gültigen Äußerung durch das automatische Spracherkennung-(ASR)-System (114, 116, 118);
 das Initialisieren einer Anforderung zur Suche in der festen kommentierten ASR-Corpusdatei (122) nach dem Auftreten der detektierten gültigen Äußerung;
 das Durchführen einer Suche mittels eines Laufzeit-Interpreter-Mittels (124) durch die Inhalte der festen kommentierten ASR-Corpusdatei (122), die im Systemspeicher geladen sind, um eine gültige Äußerung unter den aufgelisteten erwarteten gültigen Äußerungen zu finden, die der detektierten gültigen Äußerung entsprechen, und das Identifizieren der Bedeutung in der detektierten gültigen Äußerung als die Zeichendaten, die mit der gefundenen gültigen Äußerung verknüpft sind; und
 20 das Zurückgeben der Zeichendaten, die der Bedeutung in der detektierten gültigen Äußerung in der festen kommentierten ASR-Corpusdatei (122) entsprechen, an den Anforderer.

Revendications

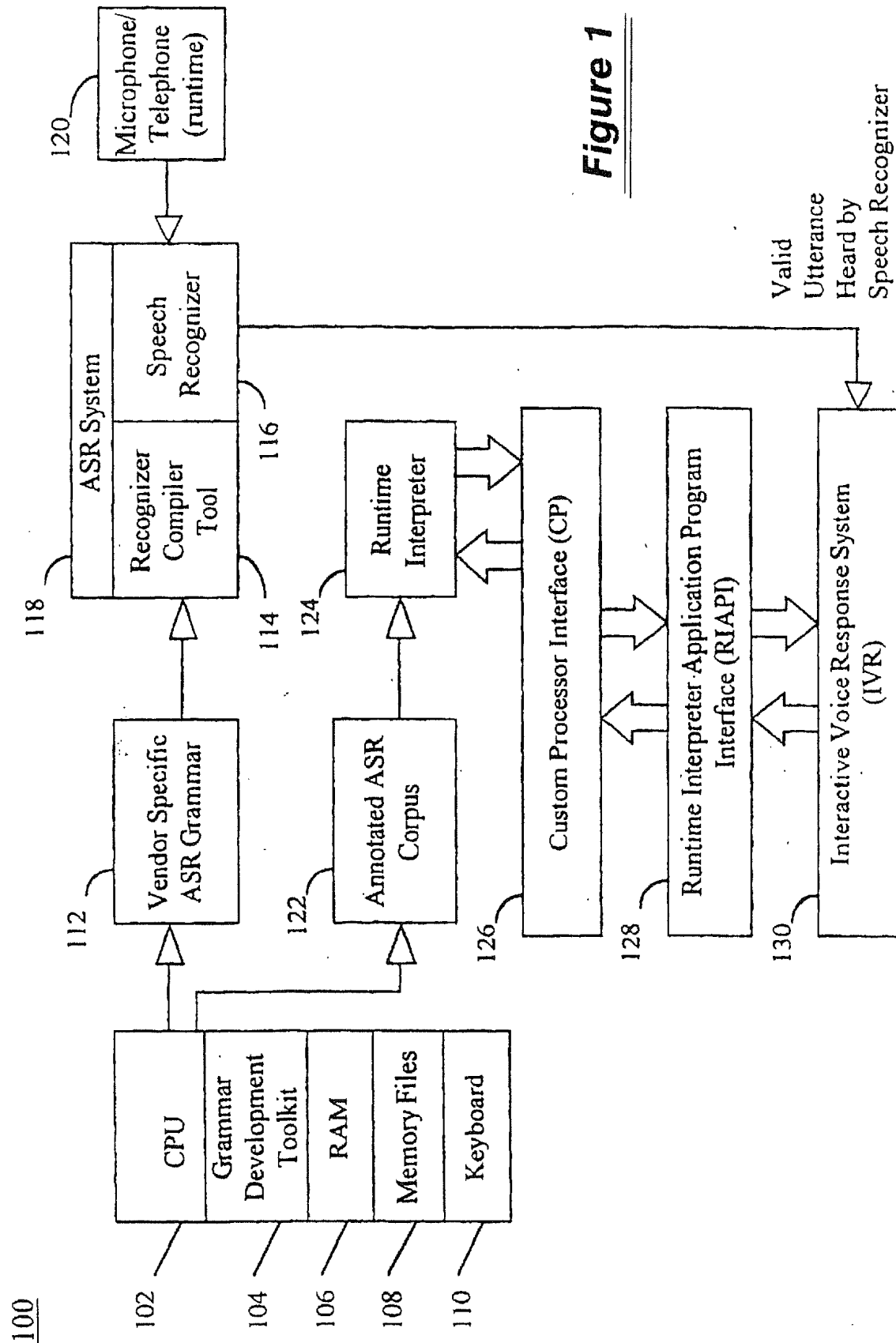
30 1. Système informatique destiné à identifier la signification représentée par une énonciation parlée valide dans une grammaire connue, ledit système comprenant :

une unité de traitement centrale (UC) (102),
 une mémoire de système reliée à ladite unité centrale (102) afin de recevoir et mémoriser des fichiers de mémorisation (108),
 35 une partie de mémoire vive (RAM) (106) de ladite mémoire de système, destinée à recevoir et à mémoriser temporairement des données durant le fonctionnement de ladite unité centrale (102),
 un fichier de corpus de reconnaissance automatique de la parole (ASR) annoté fixe prédéterminé (122), mémorisé dans ladite mémoire de système, contenant une liste de toutes les énonciations valides prévues dans ladite grammaire connue et mémorisées avec celle-ci, des données de jetons représentant la signification de chacune desdites énonciations valides prévues de la liste,
 40 un système de reconnaissance automatique de la parole (ASR) (114, 116, 118) relié à ladite unité centrale (102) afin de détecter des énonciations parlées et de générer des signaux de sortie numériques indicatifs des énonciations valides détectées,
 45 un fichier de grammaire de reconnaissance ASR spécifique à un fournisseur fixe prédéterminé (112), mémorisé dans ladite mémoire de système, relié audit système de reconnaissance ASR (114, 116, 118) et contenant des données représentant toutes les énonciations valides à détecter par ledit système de reconnaissance ASR (114, 116, 118) sous une forme qui est compatible avec le système de reconnaissance ASR (114, 116, 118), et
 50 un moyen d'interpréteur à l'exécution (124) relié à ladite unité centrale (102) et audit fichier de corpus de reconnaissance ASR annoté (122) afin d'identifier, lorsque ladite unité centrale (102) reçoit lesdits signaux de sortie numériques du système de reconnaissance ASR qui sont indicatifs d'une énonciation valide détectée, la signification représentée par ladite énonciation valide détectée en effectuant une recherche dans le contenu dudit fichier de corpus de reconnaissance ASR annoté (122), et afin de renvoyer ladite signification identifiée
 55 en temps que dites données de jeton.

2. Système selon la revendication 1, dans lequel ledit moyen d'interpréteur à l'exécution (124) comprend en outre un moyen destiné à exécuter une recherche de comparaison dans ledit fichier de corpus de reconnaissance ASR

annoté (122) dans ladite mémoire de système afin de trouver les données de jeton identifiant la signification représentée par ladite énonciation valide détectée.

- 5 3. Système selon la revendication 2, dans lequel ledit moyen d'interpréteur à l'exécution (124) comprend en outre un moyen destiné à exécuter une recherche de mise en correspondance partielle dans le contenu dudit fichier de corpus de reconnaissance ASR annoté (122) lors d'un échec de ladite recherche de comparaison, où ladite recherche partielle recherche une mise en correspondance partielle de ladite énonciation valide détectée dans ledit fichier de corpus de reconnaissance ASR annoté (122).
- 10 4. Système selon la revendication 3, dans lequel ledit moyen d'interpréteur à l'exécution (124) comprend en outre un moyen de traitement de variable destiné à traiter la partie non mise en correspondance de ladite énonciation valide détectée en tant que variable pour identifier la signification représentée par ladite partie non mise en correspondance.
- 15 5. Système selon la revendication 4, dans lequel ledit moyen de traitement de variable génère des données de variable représentant la signification de ladite partie non mise en correspondance de ladite énonciation valide détectée.
- 20 6. Système selon une ou plusieurs des revendications 1 à 5, comprenant en outre un moyen d'interface de programme d'application d'interpréteur à l'exécution (RIAPI) (128), relié audit moyen d'interpréteur à l'exécution (124), afin d'accéder audit moyen d'interpréteur à l'exécution (124).
- 25 7. Système selon la revendication 6, dans lequel ledit moyen d'interpréteur RIAPI (128) comprend en outre une interface de processeur spécialisé (CP) (126), reliée audit moyen d'interpréteur RIAPI (128) et audit moyen d'interpréteur à l'exécution (124), et à un système de-réponse vocale interactive (IVR) (130) utilisé par ledit moyen d'interpréteur RIAPI (128) afin d'accéder audit moyen d'interpréteur à l'exécution (124).
- 30 8. Système selon la revendication 7, dans lequel ledit système de réponse IVR (130) est relié audit système de reconnaissance ASR (114, 116, 118) et audit moyen d'interpréteur à l'exécution (124) afin d'effectuer une recherche dans ledit fichier de corpus de reconnaissance ASR annoté (122) pour trouver lesdites données de jeton représentant la signification représentée par ladite énonciation parlée détectée.
- 35 9. Système selon une ou plusieurs des revendications 1 à 8, où ledit système informatique comprend un système de réseau, une pluralité desdits moyens d'interpréteur à l'exécution (508a ... 508n) étant répartis sur une pluralité d'ordinateurs dans ledit réseau du système informatique.
- 40 10. Procédé destiné à identifier la signification représentée par une énonciation parlée valide dans une grammaire connue, comprenant les étapes consistant à :
45 charger un fichier de corpus de reconnaissance automatique de la parole (ASR) annoté fixe prédéterminé (122) dans une mémoire de système d'un ordinateur, où ledit fichier de corpus de reconnaissance ASR annoté fixe (122) contient une liste de toutes les énonciations valides prévues dans ladite grammaire fixe et des données de jeton représentant la signification de chacune desdites énonciations valides prévues de la liste, charger un fichier de grammaire de reconnaissance ASR spécifique à un fournisseur fixe prédéterminé (112)
50 dans ladite mémoire de système, contenant des données représentant toutes les énonciations valides à détecter par un système de reconnaissance ASR (114, 116, 118) sous une forme qui est compatible avec le système de reconnaissance ASR (114, 116, 118),
détecter une énonciation valide par ledit système de reconnaissance automatique de la parole (ASR) (114, 116, 118),
55 lancer une demande pour rechercher dans ledit fichier de corpus de reconnaissance ASR annoté fixe (122) une occurrence de ladite énonciation valide détectée,
exécuter une recherche par un moyen d'interpréteur à l'exécution (124) dans le contenu dudit fichier de corpus de reconnaissance ASR annoté fixe (122) chargé dans ladite mémoire de système pour trouver une énonciation valide parmi lesdites énonciations valides prévues de la liste correspondant à ladite énonciation valide détectée, et identifier la signification représentée par ladite énonciation valide détectée en tant que données de jeton associées à l'énonciation valide trouvée,
renvoyer lesdites données de jeton correspondant à la signification représentée par ladite énonciation valide détectée dans ledit fichier de corpus de reconnaissance ASR annoté fixe (122) vers le demandeur.

**Figure 1**

Name	Accepts	Returns
TIME1	one o'clock p.m.; thirteen hundred hours	1:00 PM (normal format)
TIME2	one o'clock p.m.; thirteen hundred hours	13:00 (European/military format)
INTEGER1	one hundred twenty three	digits
INTEGER2	one two three . . . (ten digits)	digits
CURRENCY1	twenty three dollars and fifteen cents	23.15
CURRENCY2	two three dollars and fifteen cents	23.15
DATE1	march first nineteen ninety seven; the first of march	03/01/1997 (American format)
DATE2	march first nineteen ninety seven; the first of march	01/03/1997 (European format)

Figure 2

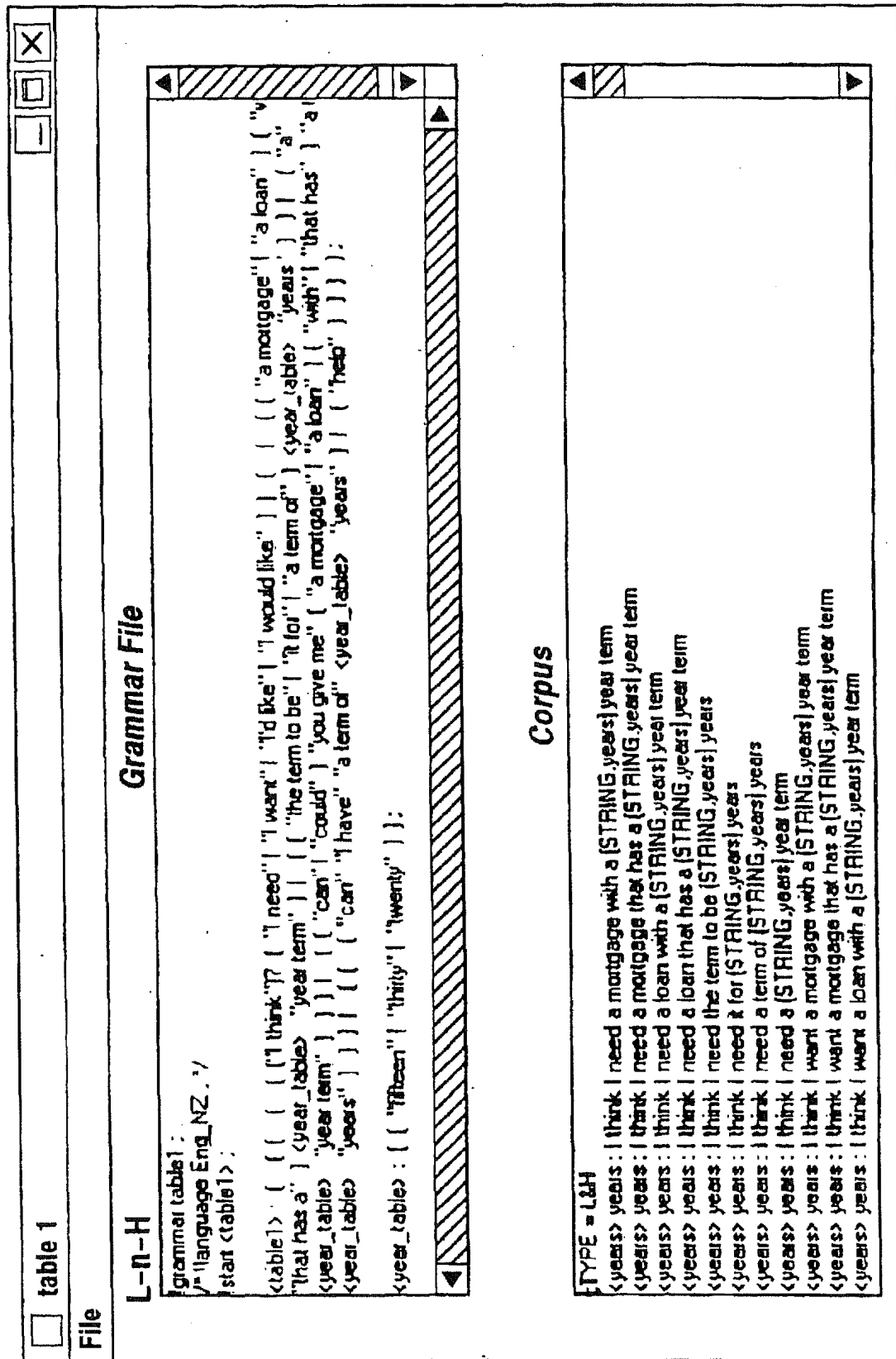
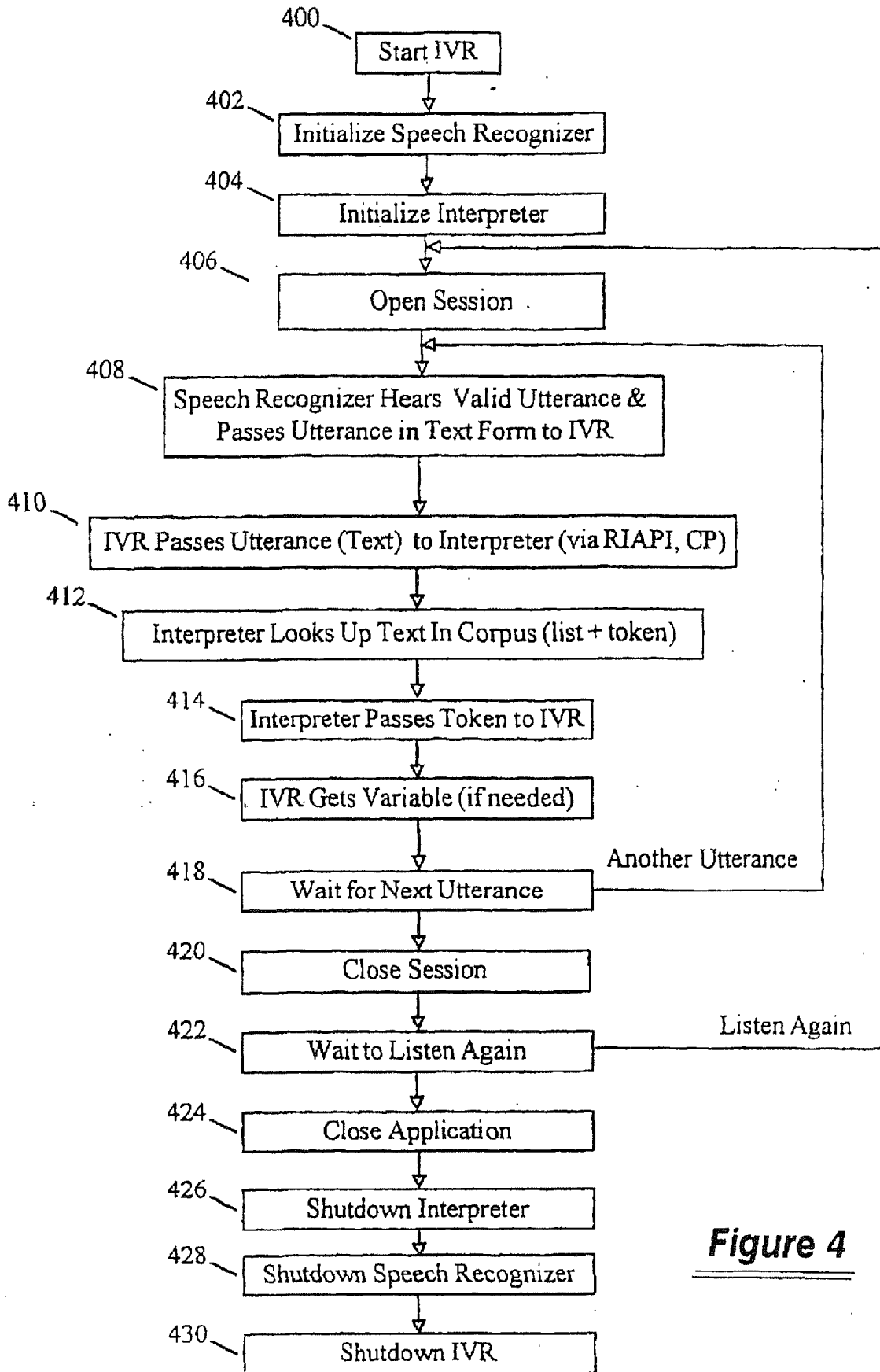
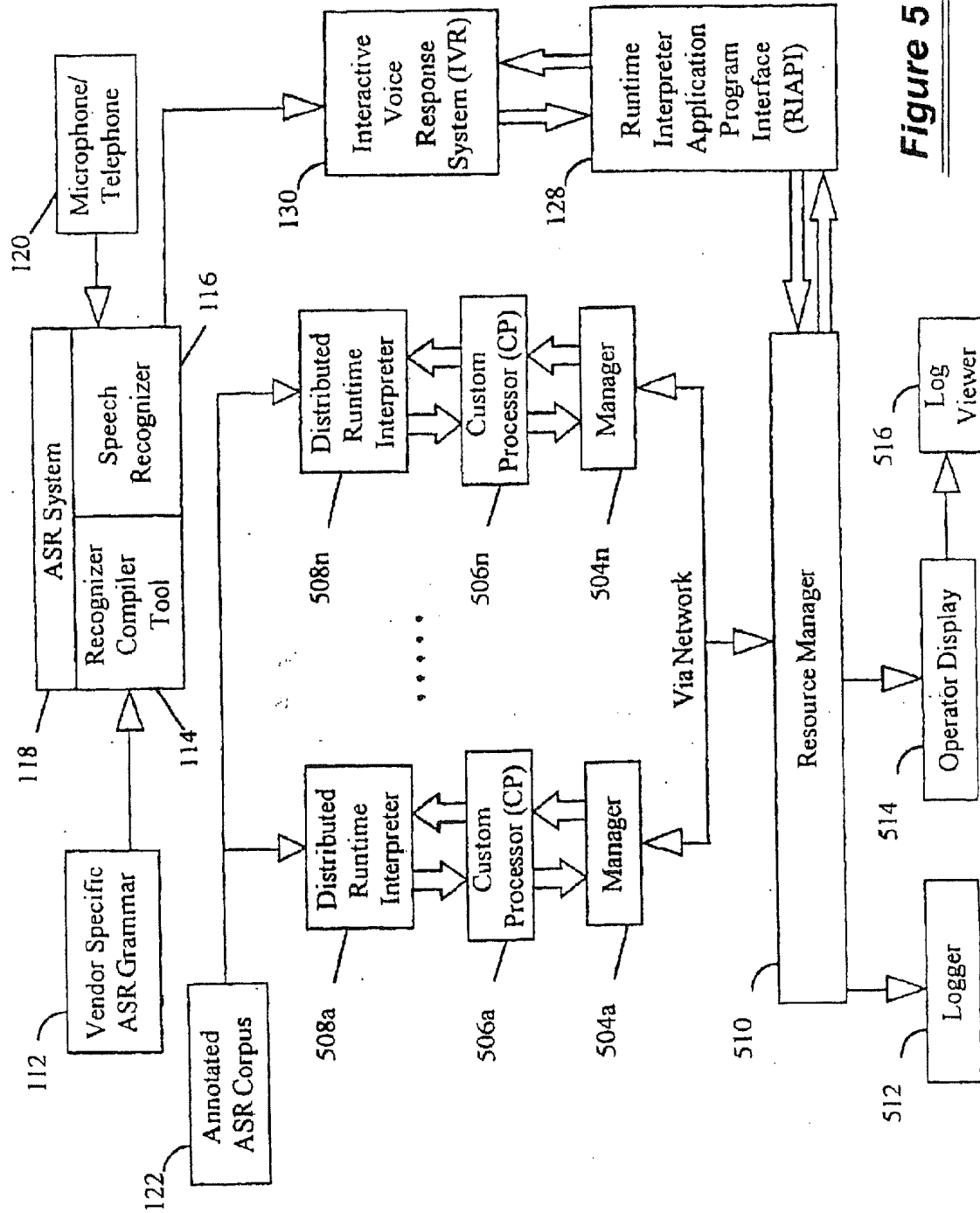


Figure 3

**Figure 4**

**Figure 5**